JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

Cite This: J. Chem. Inf. Model. 2020, 60, 982−988

pubs.acs.org/jcim

Rational Design of Chymotrypsin Inhibitor 2 by Optimizing Non-Native Interactions

Fernando B. da Silva,[†] Vinícius M. de Oliveira,[‡] Murilo N. Sanches,[†] Vinícius G. Contessoto,[¶,§] and Vitor B. P. Leite*[,†]

[†]Department of Physics, Institute of Biosciences, Humanities and Exact Sciences, São Paulo State University (UNESP), São José do Rio Preto, São Paulo 15054-000, Brazil

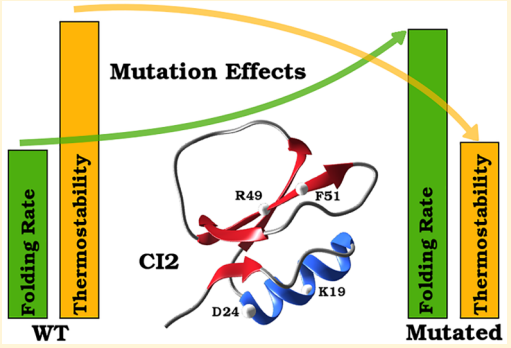
[‡]Brazilian Biosciences National Laboratory, Brazilian Center for Research in Energy and Materials, LNBio/CNPEM, Campinas, São Paulo 13083-970, Brazil

[¶]Brazilian Biorenewables National Laboratory - LNBR, Brazilian Center for Research in Energy and Materials - CNPEM, Campinas, São Paulo 13083-100, Brazil

[§]Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States

Supporting Information

ABSTRACT: Rational design of proteins via mutagenesis is crucial for several biotechnological applications. A significant challenge of the computational strategies used to predict optimized mutations is to understand the influence of each amino acid during the folding process. In the present work, chymotrypsin inhibitor 2 (CI2) and several of its designed mutants have been simulated using a non-native hydrophobic and electrostatic potential as a structure-based Cα model. Through these simulations, we could identify the most critical folding stage to accelerate CI2 and also the charged residues responsible for providing its thermostability. The replacement of ionizable residues for hydrophobic ones tended to promote the formation of the CI2 secondary structure in the early transition state, which speeds up folding. However, this same replacement destabilized the native structure, and there was a decrease in the protein thermostability. Such a simple method proved to be capable of providing valuable information about thermodynamics and kinetics of CI2 and its mutations, thus being a fast alternative to the study of rational protein design.



INTRODUCTION

The improvement of protein stability through rational mutations is a significant scientific challenge, and protein engineering has a wide range of applications in medical and industrial areas.[1−4] The development of new protein-based drugs and the use of enzymes with improved thermostability to perform catalysis in better conditions are some examples of these applications.[3−7] Such relevance in scientific and industrial scenario motivated the development of rational mutations predictors that range from force-field based molecular simulations to algorithms based on machine learning (ML).[8−14] Although there are advances in these predictor tools, issues such as data overfitting in these predictions are still recurrent, highlighting the fact that some mutation effects remain unclear.[15,16] In such a scenario, understanding the influence of each amino acid in protein folding and stability is fundamental to developing protein engineering.

Recent works by our research group have explored the addition of non-native potentials in structure-based models (SBM), which can be a powerful method to rationalize thermodynamic and kinetic effects of specific mutations in different proteins.[17−20] Using these simple models, we were able to explore the entire folding process and quantify the importance of each residue in this process using less computing power than other molecular dynamics approaches. Furthermore, even when coarse-grained models were employed, we could achieve good agreement with experimental results regarding the thermostability and folding speed.[17−20]

In our previous work, we have studied how non-native interactions modulate the protein folding rates,[20] in which hydrophobic and electrostatic potentials were added in the SBM. This kind of approach is widely studied by scientific groups to investigate different biological systems.[21−24] Therefore, proteins with mutations that might induce-or not-an outstanding performance, in terms of thermodynamics and kinetics parameters, become an attractive target for our studies. In this case, more precisely, chymotrypsin inhibitor 2 (CI2) was chosen because this protein is one of the most studied proteins. Moreover, there are many experimental results about

Special Issue: Molecular Simulation in Latin America: Coming of Age

Received: October 1, 2019
Published: December 3, 2019

DOI: 10.1021/acs.jcim.9b00911
J. Chem. Inf. Model. 2020, 60, 982−988

rational design to speed up the CI2 folding, in which single mutations decreased the protein stability by swapping charged or hydrophobic groups.

CI2 is a 65 residue monomeric protein with a single-domain, in which the tertiary structure consists of one hydrophobic α-helix and four β-sheets, see Figure 1.[27] The folding process

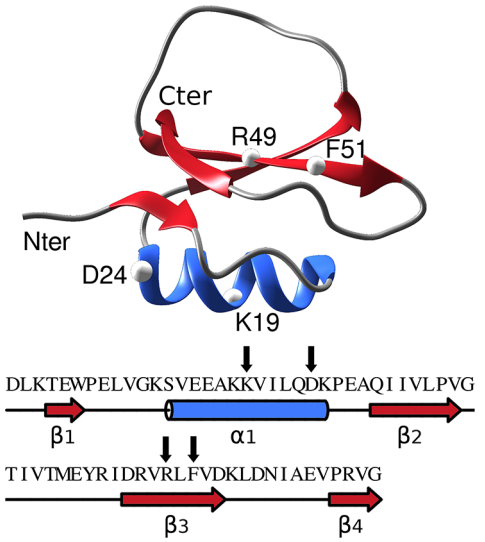


Figure 1. Three-dimensional structure of CI2 (*top*) and FASTA sequence with respective secondary structure (*bottom*), PDB ID: 2CI2. The β1, β2, β3, and β4-sheets are colored in red, and the single α1-helix is colored in blue. The residues K19, D24, R49, and F51 in the tertiary structure are the set of residues that speed up the folding time when changed by hydrophobic or charged groups. The models were drawn using CHIMERA[25] to tertiary structure and ALINE[26] to FASTA sequence and secondary structure.

occurs via two-state kinetics with a transition state, in which the α-helix is almost intact while the β-sheets are shattered.[28,29] Protein engineering studies have demonstrated that mutation effects may also increase or decrease its thermostability. Additionally, such mutations intensified local interaction in the α-helix, even eliminating some unfavorable interactions in a specific region that includes a group of arginines (three and four β-sheets). In the present study, we investigated these details with molecular dynamics simulation, using a structure-based Cα model (SBM-Cα), with hydrophobic and electrostatic non-native potentials. Mutations that accelerate folding (K19R, D24A, R49F, and F51L) were combined to provide different CI2, as shown in Figure 1. Therefore, an in-depth examination of these substitutions will shed light on questions that remain unanswered regarding cooperative effects, discussed in terms of thermostability and kinetic parameters, by rational design in protein folding.

METHODS

SBM-Cα with Non-Native Potential. The structure-based model (SBM) defines the potential in such a way that the minimum energy value is attributed to the protein's native structure. This model has been efficient in investigating kinetic and thermodynamic properties in protein folding processes.[30,31] In the SBM-Cα, the amino acids are represented by beads at the Cα carbon position. This representation considers that all Cα carbons have the same mass and radius. The Hamiltonian to SBM-Cα with non-native interactions is given by the term V_T, according to the following equation

$$V_T = V_{SBM}^{C_\alpha} + V_{ele} + V_{hp} \tag{1}$$

where the $V_{SBM}^{C_\alpha}$ is the potential to SBM-Cα (see the Supporting Information for details);[31] V_{ele} is the potential that takes into account the electrostatic interactions; and the potential that represents the hydrophobic effects is given by V_{hp}.

The electrostatic non-native interaction is taken into account by adding a charge at the Cα carbon, in which the positively charged are given by histidine (H), lysine (K), arginine (R), and glutamic acid (E), while aspartic acid (D) is negatively charged.[17−20,23,32] The interaction between the charged beads is given by

$$V_{ele} = \sum K_{ele} \frac{q_i q_j \exp(-\kappa r_{ij})}{\epsilon_K r_{ij}} \tag{2}$$

where q_i and q_j are the charges, K_{ele} = 332 kcal Å/(mol e^2), ϵ_K = 80, and κ is the inverse of Debye length.[33]

The hydrophobic potential is represented by an attractive Gaussian potential,[34] and it takes into account the interaction between two hydrophobic residues (Ala, Val, Leu, Ile, Met, Trp, and Phe) that are not in contact in the native structure

$$V_{hp} = \sum_{i}^{M} \sum_{j=i+4}^{M} K_{hp} \kappa_{ij} \exp[-(r_{ij} - \sigma)^2] \tag{3}$$

where M is the number of hydrophobic residues, K_{hp} is the strength of the hydrophobic forces, which is equal to 0.1 in this work, and κ_{ij} is the non-native hydrophobic energy given by the value of the upper triangle in Table 5 of Miyazawa and Jernigan.[35]

TKSA-MC Model. We have used the TKSA-MC Web Server in order to evaluate the electrostatic effect of each ionizable residue for the CI2 thermal stability.[10] This server uses the Tanford-Kirkwood approach with a solvent accessibility correction.[36] In a nutshell, the TKSA method consists of considering the protein as a sphere of low dielectric constant centered in its center of mass. All the charges are inside this sphere and are placed in the position of the charged atoms. Outside the sphere, the dielectric constant is equal to 78.7 (dielectric constant of water), and the salt is treated implicitly.[37−39] The charge−charge interaction of a pair of titratable residues is

$$U_{ij} = e^2 \left(\frac{A_{ij} - B_{ij}}{2b} - \frac{C_{ij}}{2a} \right) (1 - SA_{ij}) \tag{4}$$

where e is the elementary charge; A_{ij}, B_{ij}, and C_{ij} are parameters of Tanford-Kirkwood solution related to the ionizable residues position, ionic strength, and dielectric constant;[40−42] b is the radius of the protein; a is the distance of minimum approximation of ions; and SA_{ij} is the solvent accessibility surface area of a pair of charged residues.[38] Finally, the protonation states of the titratable residues are sampled by the Metropolis Monte Carlo method.[43]

The Probability of Contact Formation. The probability of formation of a native contact, P_{ij}, has been calculated in the folding process. The formation of two native contacts, i and j, in the transition state (TS) is computed in relation to unfolded and folded states. Thus, P_{ij} is a statistical parameter calculated

DOI: 10.1021/acs.jcim.9b00911
J. Chem. Inf. Model. 2020, 60, 982−988

from the folding dynamics trajectory. The ΔP_{ij} is the difference between two native contact maps. In this work, we present the percentage difference of them. Therefore, when a set of residues is swapped with other amino acid groups, the ΔP_{ij} reflects the change.

Simulation Details. All simulations were carried out using the molecular dynamics package Gromacs (version 4.5.5), with the input files obtained from the SMOG webtool,[44] and the contact map of the protein defined by the Shadow Contact Map.[45] The time step was set at 0.5 fs, and the simulation was performed over 5×10^9 steps, recording snapshots every 5000 steps with an initial random unfolded structure. To maintain the simulation coupled with a thermal bath, the Berendsen thermostat algorithm was employed, with a constant equal to 1 ps.[46]

Analysis. To follow the folding process of the simulation, it is necessary to choose a reaction coordinate, which in this case was the fraction of native contacts (Q). Two native contacts, i and j, are formed when the distance between them is shorter than $1.2d_{ij}$. d_{ij} is the distance between two C_α's in the native structure with $j > i + 3$. The thermodynamics properties calculation was performed by merging multiple simulations carried out under constant temperature and using the weighted histogram analysis method (WHAM).[47] Following the experimental conditions, the mean first passage time (MFPT) was simulated with all proteins at the same temperature of the wild-type. Over 1000 independent simulations were computed to calculate the folding time τ as an average of the folding time of each simulation, that started with an open random structure ($Q_{unf} \approx 0.1$) and was executed until it reached the folded state ($Q_{fold} \approx 0.8$).

RESULTS AND DISCUSSION

The CI2 Mutants. In this work, we have performed four CI2 versions with the mutants that accelerate folding (K19R, D24A, and R49F/F51L). All newly engineered CI2s (M_A, M_B, M_C, and M_D) were designed as a combination of mutants above.[48,49] These mutants were generated by swapping charged groups to others, positively or negatively, as well as charged groups by hydrophobic ones and vice versa. Thus, M_A (K19R/D24A), M_B (K19R/R49F/F51L), M_C (D24A/R49F/F51L), and MD (K19R/D24A/R49F/F51L) native structures were generated by MODELLER software.[50]

Electrostatic Interaction and Stability. Among the several interactions responsible for providing protein thermostability, the electrostatic effect is known to be critical.[17−19,24,51] Using the TKSA-MC tool, we were able to compute the electrostatic free energy contribution of each ionizable residue of the protein, ΔG_{qq}. In Figure 2, we present these electrostatic free energy profiles of CI2 wild-type and its mutated form M_D. All the other mutant's electrostatic profiles are presented in the SI.

The CI2 wild-type is already a very electrostatically stable protein, with almost all ionizable residues contributing to stabilizing its native structure, and only D53 and D56 having a significant positive electrostatic contribution. The summation of the electrostatic contribution of all residues is the ΔG_T, and the more negative this value, the more stable the protein tends to be. Consequently, we expected the folding temperature to increase. The mutated protein M_D has the highest ΔG_T among all proteins tested in this work and is near 11.0 kJ/mol more positive than CI2 wild-type. Such an increase in electrostatic free energy is caused by the mutation of three charged residues

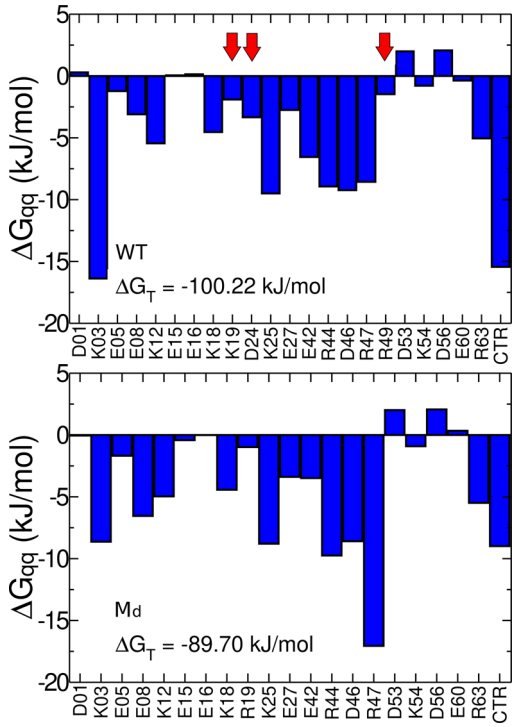


Figure 2. (*top* and *bottom*) Electrostatic free energy profiles of each ionizable residue of wild-type (WT) and engineered M_D - CI2 (K19R/D24A/R49F/F51L), respectively. $\Delta G_{qq} > 0$ are the residues with positive electrostatic free energy and SASA greater than 50%. All the results were obtained in a pH of 7.0. The charged residues that will be mutated are indicated by arrows in red in the *top* part.

K19R, D24A, and R49F, which were contributing negatively to the ΔG_T wild-type, for hydrophobic ones (the mutated residues are marked by red arrows in Figure 2).

In Figure 3A, we present the values of the electrostatic free energy as a function of CI2 wild-type and mutants folding temperature, T_f. Due to the lack of information about experimental data, the analyses regarding the electrostatic free energy versus T_f are not provided in this work. For that reason, only theoretical and computational results are shown in Figure 3A, where the mutations proposed and studied experimentally are in orange color, and the mutations introduced as a model in this work are in blue color. The T_f of each protein was determined by their peaks of heat capacity curves, which were obtained applying the WHAM in several simulations of SBM-Cα with non-native potential (see Simulation Details). The increase of ΔG_T tends to promote a decrease in T_f as observed in our results (correlation of −0.91). It should be noted that the TKSA-MC takes into account only the electrostatic effect in the native structure of the protein, whereas our simulations take into account two non-native interactions, hydrophobic and electrostatic, calculated in the entire folding process. The remarkable correlation between ΔG_T and T_f indicates that the electrostatic interaction is the driving force in the thermostability of CI2 and that the replacement of charged amino acids for hydrophobic ones causes protein destabilization.

Kinetic Effects by Rational Design. Data from ref 48 revealed that the K19 and D24 residues, located in C-terminus of a single α-helix, are essential in preserving the stability of the whole native structure, but the D24 residue presents the main destabilizing effect on an isolated α-helix. For that reason, they

DOI: 10.1021/acs.jcim.9b00911
J. Chem. Inf. Model. 2020, 60, 982−988

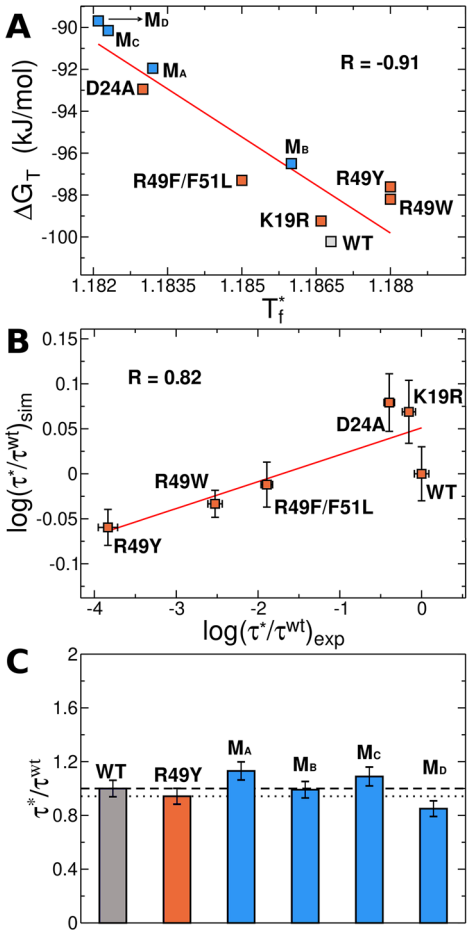


Figure 3. (A) The electrostatic free energy (ΔG_T) of each engineered CI2 version (M_A, M_B, M_C, and M_D) and wild-type (WT) as a function of folding temperature (T_f^*) in reduced units. (B) The mutation's effects in the folding time of the engineered CI2 version are normalized by its wild-type and are represented by its folding time ratio logarithm, calculated to simulation results and experimental data. The linear correlation is presented as a solid line in (A) and (B). (C) Engineering CI2 folding times, τ^*, normalized by CI2 WT, τ^{wt}. The blue and gray bars are the computational results to new mutants, and the orange bar is the folding time ratio for experimental data. All of the experimental data in (B) and (C) is taken from refs 48 and 49.

have mutated K19 and D24 using arginine (R) and alanine (A), respectively. The results of these mutations preserved the helix formation and accelerated its folding time. On the other hand, they decrease the stability of the whole protein. Another critical region in CI2 wild-type includes several arginines (R47, R49, and R63), in which their arrangement promotes an electrostatic repulsion between their charged groups. However, Ladurner et al.[48] and Lawrence et al.[49] have shown in their experimental studies that this region is highly conserved. Likewise, the mutations on the R49 by large hydrophobic groups lead to a notable folding time reduction.[48,49] Furthermore, R49 forms relevant interactions with other residues, particularly with F51 in TS.

We investigated the kinetics parameters focusing on folding time for the CI2 wild-type and the K19R, D24A, R49F/F51L, R49W, and R49Y mutations, using non-native hydrophobic and electrostatic potential in the SBM-Cα. In Figure 3B, we show the logarithm folding time ratio for both experimental and computational data. The comparison between the two reveals a linear correlation of 0.82, and this result suggests that our simulation approach is in qualitative agreement with experiments, despite two outlier points: WT and K19R.

According to the results in the literature, we have taken only the best accelerated CI2 version, R49Y, to compare with the proposed mutants in this work, Figure 3C. Folding time ratio results show that the proposed mutations, M_A and M_C, do not speed up the folding but, rather, decelerate it. Unlike M_A and M_C results, the mutation M_B presented almost the same folding time as CI2 wild-type. However, the folding time calculation of M_D presented to be the fastest of all the mutations designed. Furthermore, the M_D has also shown to be more accelerated than wild-type and R49Y. The set of mutations that we have studied here gave us some insights on how to combine mutations that already speed up the folding to accelerate it even more. We have also noticed in this set of proposed CI2 versions a lower thermostability promoted by an increase in the electrostatic free energy. Based on this assumption, we calculated some parameters to further analyze the electrostatic and hydrophobic effects of M_D (see the next section).

Antagonistic Issues on Engineering CI2. The CI2 with the mutation M_D is the fastest engineered CI2 version that we have designed (see Figures 2 and 3). As discussed above, we observed that it presents a folding time ratio lower than the WT CI2 and the R49Y. It is worth mentioning that the SBM-Cα takes into account only the native interaction contribution to protein folding, but some limitations in this model are noted to study mutation effects. For this reason, we added two non-native potentials to SBM-Cα that can elucidate some details. On that account, the methodology applied here allows us to observe the effect of changing charged amino acid groups in terms of hydrophobic and electrostatic energy. The hydrophobic and electrostatic energies have been calculated in different stages to the WT and engineered CI2M_D, see Figure 4A. We have evaluated these energy components for the unfolded state (Unf - $19 < Q < 38$), the early transition state (ETS - $39 < Q < 56$), the transition state (TS - $57 < Q < 82$), the late transition state (LTS - $83 < Q < 112$), and the folded state (Fold - $113 < Q < 141$).

The WT has shown to be more electrostatically favorable than the M_D in all stages. On the other hand, the M_D was observed to be more hydrophobically favorable than the WT in all stages. The important feature is the overall free energy compared along all stages; M_D is less stable than the WT, $\Delta\Delta G_T = -10$ kJ/mol, with T_{fs}, 1.182 and 1.187 in reduced units, respectively, which corresponds to just 0.6% difference, as shown in Figure 3A. On the other hand, the M_D free energy barrier is smaller than the WT by just 0.4 kT, but the combination of effects yields to a folding time 20% smaller, see Figure 3C. The electrostatic energy in the transition from ETS to LTS tends to increase in both proteins. For WT the hydrophobic energy increases monotonically from ETS to the folded state, while for the M_D $\langle E^{hp} \rangle$ oscillates along the folding process. These different behaviors of the WT and M_D in the transition between specific stages led us to calculate the contact map. To investigate the main effects of designed CI2M_D to speed up the protein folding, we calculated the native contact map to ETS and TS.

In Figure 4B, we calculated the difference of native contact formation probability (ΔP_{ij}) between MD and wild-type. In the ETS (upper triangle), the M_D mutant presents native contacts in red color with higher probability formation than WT. In this case, this region corresponds to the same place

DOI: 10.1021/acs.jcim.9b00911
J. Chem. Inf. Model. 2020, 60, 982−988

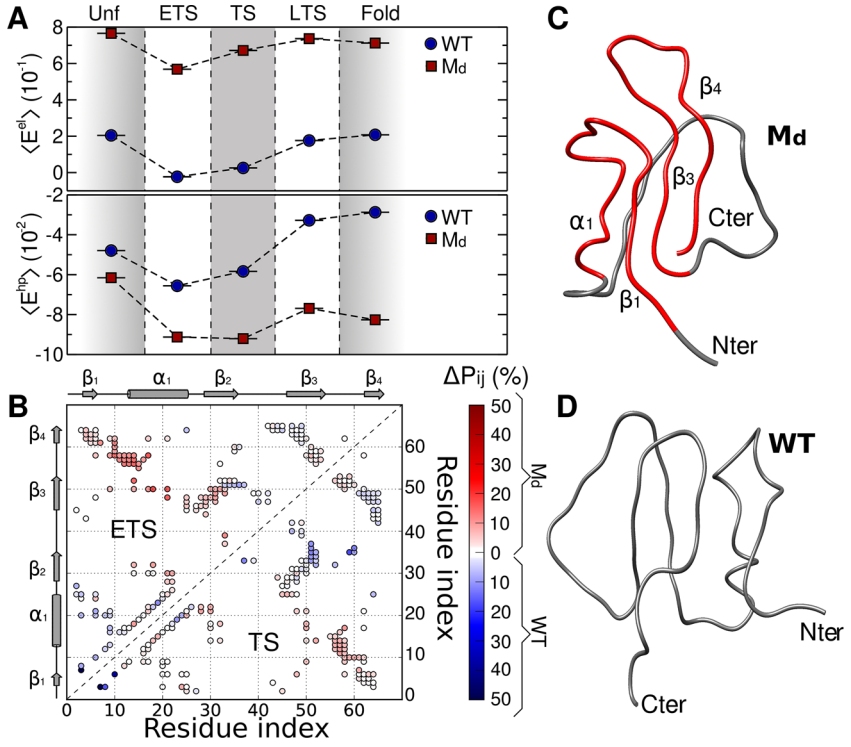


Figure 4. (A) The average of the electrostatic ($\langle E^{el} \rangle$) and hydrophobic ($\langle E^{hp} \rangle$) energy to CI2 wild-type (WT - blue) and mutant (M_D - red), respectively. The energy calculated for the unfolded state (Unf), early transition state (ETS), transition state (TS), late transition state (LTS), and folded state (Fold) to both proteins. (B) The upper and lower triangles are the contact map difference between the mutant (M_D) and the WT in the early transition state (ETS - upper triangle) and in the transition state (TS - lower triangle), respectively. The secondary structures are represented in gray color, and the color bar is associated with the difference of contact map probabilities (ΔP_{ij} (%)) between CI2 (MD) and WT. (C) and (D) are examples of two representative structures of M_D and WT in the early transition state. The highlighted region in the M_D corresponds to the same region in red in the contact map (B) of the ETS part.

that the mutations have performed, α-helix one (α_1) and β-sheet three (β_3). In blue scale, we have the representation of the native contacts formed preferred in the wild-type, and in white are the native contacts with the same probability ($\Delta P_{ij} \cong 0$); that is, the probability of contact formation in the wild-type or in the mutant is approximately the same. In the lower triangle, we have the native contact map difference between M_D and wild-type in the TS stage. In this specific case, we noticed that both proteins have shown almost the same native contact formation probability, and consequently, they do not present significant differences in folded regions in TS, unlike what occurs in ETS. Therefore, our results indicate that ETS is a prime stage to speed up the MD folding. Hence, its structures were analyzed in this stage and compared with CI2 WT.

The main representative structures for the WT and engineered CI2M_D in the ETS stage are presented in Figure 4C and 4D, respectively. According to the highlighted red region in the native contact map, we highlighted the representative structure of M_D. This structure displays an unstructured α-helix, while the β-sheet one, three, and four tend to be paired, as in the native structure (see Figure 1). In contrast to what was observed for the M_D, the significant representative structure of wild-type in the ETS also presents an unstructured α-helix, but its β-sheets are not paired. The N and C terminus of wild-type are pulling away, while the N and C terminus of M_D are more compacted with the rest of its structure. According to experimental data from ref 48, the removal of some electrostatic interactions increases the packing of protein, even while decreasing the folding time. Our computational results are in a reasonable agreement with experimental data and an energy decrease was observed in the M_D, as well as a favorable stage in the protein folding determined by simulation in which is described one of the main reasons that the folding time speeds up. From this standpoint, the replacement of charged residues by hydrophobic ones not only decreases the folding time but also is responsible for making the engineered CI2M_D more compact in the ETS than the wild-type.

The addition of non-native interactions in different systems showed similar results as those presented here. Those interactions can speed up the folding by decreasing the free energy barrier.[52,53] The optimum amount of non-native interactions in a protein depends on its structure and topology.[54] Any value beyond the optimum non-native interactions value leads a protein to a slower folding process by increasing the number of local minima in the ETS.[55] This concept is related to frustration in proteins, and the mutation M_D is an example of such behavior. The removal of some electrostatic interactions may decrease frustration and limit competition between native interactions and non-native interactions.

■ CONCLUSION

In-depth knowledge of the effect of each amino acid in protein folding and stability is fundamental to the development of protein engineering. In the present work, we have used computational methods that combine electrostatic and hydrophobic interactions with structure-based models to study the effects of several mutations in the kinetics and thermodynamics

DOI: 10.1021/acs.jcim.9b00911
J. Chem. Inf. Model. 2020, 60, 982−988

of CI2 protein. Such a method has already been tested in a recent work by our research group and presented new insights about cooperative non-native interactions in the folding of spectrin domains, in which we observed a cooperative non-native interaction (electrostatic and hydrophobic contributions) for the folding time of spectrin domains (R15, R16, and R17).[20] In the present case, we noticed an uncommon behavior for the WT and the M_D, in which some of the proposed mutations on CI2 do not present a non-native cooperative effect but an opposite effect promoting antagonistic issues in thermostability and folding speed. The engineered mutation MD is the primary example of this effect: while its mutations cause a significant increase in its folding speed, the same mutations cause a severe decrease in its thermostability. This effect is counterintuitive, as mutations that improve thermostability are usually associated with the acceleration of the folding process.[24,56] Our results also show that the replacement of charged residues with hydrophobic ones tends to promote the formation of secondary structures already in the ETS stage of folding, and it accelerates the CI2 folding process. This replacement also changes the electrostatic energy profile of the native state of the protein, which causes the loss of thermostability. Therefore, the simple methodology used here has provided valuable information about mutation effects and identified the most critical stages of folding as they relate to thermostability and folding speed.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.9b00911.

Additional SBM-Cα details and figure with electrostatic free energy, ΔG_T, to K19R, R49W, R49Y, R49F/F51L, M_B (K19R/R49F/F51L), D24A, M_A (K19R/D24A), M_C (D24A/R49F/F51L) mutants, respectively (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: vitor.leite@unesp.br.

ORCID

Fernando B. da Silva: 0000-0002-0285-8700
Vinícius M. de Oliveira: 0000-0003-0927-3825
Vinícius G. Contessoto: 0000-0002-1891-9563
Vitor B. P. Leite: 0000-0003-0008-9079

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

F.B.S. was supported by the National Council for Scientific and Technological Development (CNPq - Grant Process No. 141715/2017-0). V.M.O. was supported by the São Paulo Research Foundation (FAPESP), Grant No. 2018/11614-3. M.N.S. was supported by the São Paulo Research Foundation (FAPESP) Grant No. 2017/25130-5. V.G.C. is a Robert A. Welch Postdoctoral Fellow and was also funded by Grants 2016/13998-8 and 2017/09662-7, FAPESP (São Paulo Research Foundation) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. V.B.P.L. was supported by the São Paulo Research Foundation (FAPESP), Grants 2018/18668-1 and 2016/19766-1. This work was partially supported by the Center for Theoretical Biological Physics sponsored by the NSF (Grant PHY- 1427654) and by NSF-CHE 1614101. We also thank the Center for Scientific Computing (NCC/Grid-UNESP) of São Paulo State University (UNESP) for computational resources.

■ REFERENCES

(1) Yang, H.; Liu, L.; Li, J.; Chen, J.; Du, G. Rational Design to Improve Protein Thermostability: Recent Advances and Prospects. *ChemBioEng Rev.* **2015**, *2*, 87−94.

(2) Seeliger, D.; Schulz, P.; Litzenburger, T.; Spitz, J.; Hoerer, S.; Blech, M.; Enenkel, B.; Studts, J. M.; Garidel, P.; Karow, A. R. Boosting Antibody Developability Through Rational Sequence Optimization. *mAbs* **2015**, *7*, 505−515.

(3) Feng, R.; Liang, B.; Hou, C.; Han, D.; Han, L.; Lang, Q.; Liu, A.; Han, L. Rational Design of Xylose Dehydrogenase for Improved Thermostability and Its Application in Development of Efficient Enzymatic Biofuel Cell. *Enzyme Microb. Technol.* **2016**, *84*, 78−85.

(4) Thompson, P.; Fleming, R.; Bezabeh, B.; Huang, F.; Mao, S.; Chen, C.; Harper, J.; Zhong, H.; Gao, X.; Yu, X.-Q. Rational Design, Biophysical and Biological Characterization of Site-Specific Antibody-Tubulysin Conjugates with Improved Stability, Efficacy and Pharmacokinetics. *J. Controlled Release* **2016**, *236*, 100−116.

(5) Galdiero, S.; Gomes, P. A. C. Peptide-Based Drugs and Drug Delivery Systems. *Molecules* **2017**, *22*, 2185.

(6) Sachdeva, S.; Lobo, S.; Goswami, T. What is the Future of Noninvasive Routes for Protein- and Peptide-Based Drugs? *Ther. Delivery* **2016**, *7*, 355−357.

(7) Mohandesi, N.; Haghbeen, K.; Ranaei, O.; Arab, S. S.; Hassani, S. Catalytic Efficiency and Thermostability Improvement of Suc2 Invertase Through Rational Site-Directed Mutagenesis. *Enzyme Microb. Technol.* **2017**, *96*, 14−22.

(8) Kellogg, E. H.; Leaver-Fay, A.; Baker, D. Role of Conformational Sampling in Computing Mutation-Induced Changes in Protein Structure and Stability. *Proteins: Struct., Funct., Genet.* **2011**, *79*, 830−838.

(9) Yin, S.; Ding, F.; Dokholyan, N. V. Modeling Backbone Flexibility Improves Protein Stability Estimation. *Structure* **2007**, *15*, 1567−1576.

(10) Contessoto, V. G.; de Oliveira, V. M.; Fernandes, B. R.; Slade, G. G.; Leite, V. B. P. TKSA-MCA: Web Server for Rational Mutation Through the Optimization of Protein Charge Interactions. *Proteins: Struct., Funct., Genet.* **2018**, *86*, 1184−1188.

(11) Masso, M.; Vaisman, I. I. Accurate Prediction of Stability Changes in Protein Mutants by Combining Machine Learning with Structure Based Computational Mutagenesis. *Bioinformatics* **2008**, *24*, 2002−2009.

(12) Huang, L.-T.; Gromiha, M. M. Reliable Prediction of Protein Thermostability Change upon Double Mutation from Amino Acid Sequence. *Bioinformatics* **2009**, *25*, 2181−2187.

(13) Capriotti, E.; Fariselli, P.; Casadio, R. I-Mutant2. 0: Predicting Stability Changes Upon Mutation from the Protein Sequence or Structure. *Nucleic Acids Res.* **2005**, *33*, W306−W310.

(14) Cheng, J.; Randall, A. Z.; Sweredoski, M. J.; Baldi, P. SCRATCH: A Protein Structure and Structural Feature Prediction Server. *Nucleic Acids Res.* **2005**, *33*, W72−W76.

(15) Jia, L.; Yarlagadda, R.; Reed, C. C. Structure Based Thermostability Prediction Models for Protein Single Point Mutations with Machine Learning Tools. *PLoS One* **2015**, *10*, No. e0138022.

(16) Fang, J. A Critical Review of Five Machine Learning-Based Algorithms for Predicting Protein Stability Changes Upon Mutation. *Briefings Bioinf.* **2019**, *00*, bbz071.

(17) Contessoto, V. G.; de Oliveira, V. M.; de Carvalho, S. J.; Oliveira, L. C.; Leite, V. B. P. NTL9 Folding at Constant pH: The Importance of Electrostatic Interaction and pH Dependence. *J. Chem. Theory Comput.* **2016**, *12*, 3270−3277.

DOI: 10.1021/acs.jcim.9b00911
J. Chem. Inf. Model. 2020, 60, 982−988

(18) Coronado, M. A.; Caruso, I. P.; De Oliveira, V. M.; Contessoto, V. G.; Leite, V. B. P.; Kawai, L. A.; Arni, R. K.; Eberle, R. J. Cold Shock Protein A from Corynebacterium Pseudotuberculosis: Role of Electrostatic Forces in the Stability of the Secondary Structure. *Protein Pept. Lett.* **2017**, *24*, 358−367.
(19) Martins de Oliveira, V.; de Godoi Contessoto, V.; da Silva, F. B.; Caetano, D. L. Z.; de Carvalho, S. J.; Leite, V. B. P. Effects of pH and Salt Concentration on Stability of a Protein G Variant Using Coarse-Grained Models. *Biophys. J.* **2018**, *114*, 65−75.
(20) Bruno da Silva, F.; Contessoto, V. G.; de Oliveira, V. M.; Clarke, J.; Leite, V. B. P. Non-Native Cooperative Interactions Modulate Protein Folding Rates. *J. Phys. Chem. B* **2018**, *122*, 10817−10824.
(21) Chen, T.; Chan, H. S. Native Contact Density and Nonnative Hydrophobic Effects in the Folding of Bacterial Immunity Proteins. *PLoS Comput. Biol.* **2015**, *11*, e1004260.
(22) Wu, J.; Chen, G.; Zhang, Z.; Zhang, P.; Chen, T. The Low Populated Folding Intermediate of a Mutant of the Fyn SH3 Domain Identified by a Simple Model. *Phys. Chem. Chem. Phys.* **2017**, *19*, 22321−22328.
(23) Yrazu, F. M.; Pinamonti, G.; Clementi, C. The Effect of Electrostatic Interactions on the Folding Kinetics of a 3-α-Helical Bundle Protein Family. *J. Phys. Chem. B* **2018**, *122*, 11800−11806.
(24) Tzul, F. O.; Schweiker, K. L.; Makhatadze, G. I. Modulation of Folding Energy Landscape by Charge−Charge Interactions: Linking Experiments with Computational Modeling. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E259−E266.
(25) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera−A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605−1612.
(26) Bond, C. S.; Schüttelkopf, A. W. *ALINE*: A WYSIWYG Protein-Sequence Alignment Editor for Publication-Quality Alignments. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2009**, *65*, 510−512.
(27) McPhalen, C.; James, M. Crystal and Molecular Structure of the Serine Proteinase Inhibitor CI-2 from Barley Seeds. *Biochemistry* **1987**, *26*, 261−269.
(28) Itzhaki, L. S.; Otzen, D. E.; Fersht, A. R. The Structure of the Transition State for Folding of Chymotrypsin Inhibitor 2 Analysed by Protein Engineering Methods: Evidence for a Nucleation-Condensation Mechanism for Protein Folding. *J. Mol. Biol.* **1995**, *254*, 260−288.
(29) Jackson, S. E.; Fersht, A. R. Folding of Chymotrypsin Inhibitor 2. 1. Evidence for a two-state Transition. *Biochemistry* **1991**, *30*, 10428−10435.
(30) Taketomi, H.; Ueda, Y.; Gō, N. Studies on Protein Folding, Unfolding and Fluctuations by Computer Simulation: I. The Effect of Specific Amino Acid Sequence Represented by Specific Inter-Unit Interactions. *Int. J. Pept. Protein Res.* **1975**, *7*, 445−459.
(31) Clementi, C.; Nymeyer, H.; Onuchic, J. N. Topological and Energetic Factors: What Determines the Structural Details of the Transition State Ensemble and "En-Route" Intermediates for Protein Folding? An Investigation for Small Globular Proteins. *J. Mol. Biol.* **2000**, *298*, 937−953.
(32) Azia, A.; Levy, Y. Nonnative Electrostatic Interactions Can Modulate Protein Folding: Molecular Dynamics with a Grain of Salt. *J. Mol. Biol.* **2009**, *393*, 527−542.
(33) Ullner, M.; Woodward, C. E.; Jönsson, B. A Debye-Hückel Theory for Electrostatic Interactions in Proteins. *J. Chem. Phys.* **1996**, *105*, 2056−2065.
(34) Chen, T.; Song, J.; Chan, H. S. Theoretical Perspectives on Nonnative Interactions and Intrinsic Disorder in Protein Folding and Binding. *Curr. Opin. Struct. Biol.* **2015**, *30*, 32−42.
(35) Miyazawa, S.; Jernigan, R. L. Self-Consistent Estimation of Inter-Residue Protein Contact Energies Based on an Equilibrium Mixture Approximation of Residues. *Proteins: Struct., Funct., Genet.* **1999**, *34*, 49−68.
(36) Havranek, J. J.; Harbury, P. B. Tanford−Kirkwood Electrostatics for Protein Modeling. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 11145−11150.
(37) Shire, S.; Hanania, G.; Gurd, F. N. Electrostatic Effects in Myoglobin. Hydrogen Ion Equilibriums in Sperm Whale Ferrimyoglobin. *Biochemistry* **1974**, *13*, 2967−2974.
(38) Tanford, C.; Roxby, R. Interpretation of Protein Titration Curves. Application to Lysozyme. *Biochemistry* **1972**, *11*, 2192−2198.
(39) Orttung, W. H. Proton Binding and Dipole Moment of Hemoglobin. Refined Calculations. *Biochemistry* **1970**, *9*, 2394−2402.
(40) Tanford, C.; Kirkwood, J. G. Theory of Protein Titration Curves. I. General Equations for Impenetrable Spheres. *J. Am. Chem. Soc.* **1957**, *79*, 5333−5339.
(41) Kirkwood, J. G. Theory of Solutions of Molecules Containing Widely Separated Charges with Special Application to Zwitterions. *J. Chem. Phys.* **1934**, *2*, 351−361.
(42) Da Silva, F. L. B.; Jönsson, B.; Penfold, R. A Critical Investigation of the TanfordKirkwood Chceme by means of Monte Carlo simulations. *Protein Sci.* **2001**, *10*, 1415−1425.
(43) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, *21*, 1087−1092.
(44) Noel, J. K.; Levi, M.; Raghunathan, M.; Lammert, H.; Hayes, R. L.; Onuchic, J. N.; Whitford, P. C. SMOG 2: A Versatile Software Package for Generating Structure-Based Models. *PLoS Comput. Biol.* **2016**, *12*, e1004794.
(45) Noel, J. K.; Whitford, P. C.; Onuchic, J. N. The Shadow Map: A General Contact Definition for Capturing the Dynamics of Biomolecular Folding and Function. *J. Phys. Chem. B* **2012**, *116*, 8692−8702.
(46) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(47) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(48) Ladurner, A. G.; Itzhaki, L. S.; Daggett, V.; Fersht, A. R. Synergy Between Simulation and Experiment in Describing the Energy Landscape of Protein Folding. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 8473−8478.
(49) Lawrence, C.; Kuge, J.; Ahmad, K.; Plaxco, K. W. Investigation of an Anomalously Accelerating Substitution in the Folding of a Prototypical two-state Protein. *J. Mol. Biol.* **2010**, *403*, 446−458.
(50) Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinf.* **2016**, *54*, 5.6.1−5.6.37.
(51) Gribenko, A. V.; Patel, M. M.; Liu, J.; McCallum, S. A.; Wang, C.; Makhatadze, G. I. Rational Stabilization of Enzymes by Computational Redesign of Surface Charge− Charge Interactions. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 2601−2606.
(52) Plotkin, S. S. Speeding Protein Folding Beyond the G¯o Model: How a Little Frustration Sometimes Helps. *Proteins: Struct., Funct., Genet.* **2001**, *45*, 337−345.
(53) Clementi, C.; Plotkin, S. S. The Effects of Nonnative Interactions on Protein Folding Rates: Theory and Simulation. *Protein Sci.* **2004**, *13*, 1750−1766.
(54) Contessoto, V. G.; Lima, D. T.; Oliveira, R. J.; Bruni, A. T.; Chahine, J.; Leite, V. B. P. Analyzing the Effect of Homogeneous Frustration in Protein Folding. *Proteins: Struct., Funct., Genet.* **2013**, *81*, 1727−1737.
(55) Mouro, P. R.; de Godoi Contessoto, V.; Chahine, J.; Junio de Oliveira, R.; Pereira Leite, V. B. Quantifying Nonnative Interactions in the Protein-Folding Free-Energy Landscape. *Biophys. J.* **2016**, *111*, 287−293.
(56) Zarrine-Afsar, A.; Zhang, Z.; Schweiker, K. L.; Makhatadze, G. I.; Davidson, A. R.; Chan, H. S. Kinetic Consequences of Native State Optimization of Surface-Exposed Electrostatic Interactions in the Fyn SH3 Domain. *Proteins: Struct., Funct., Genet.* **2012**, *80*, 858−870.

DOI: 10.1021/acs.jcim.9b00911
J. Chem. Inf. Model. 2020, 60, 982−988